

補助事業番号 2021M-227  
 補助事業名 2021年度生活習慣病危険因子検出及び注意喚起システムの開発 補助事業  
 補助事業者名 国立大学法人 徳島大学 松本 和幸

### 1 研究の概要

本研究では、普段の生活習慣に起因する生活習慣病のリスク軽減を目指し、早期の段階で生活習慣の改善を促すことを目標として、生活習慣病の危険因子を検出するアルゴリズムを開発し、そのアルゴリズムを元に、リスクの高いユーザをSNSなどから抽出する技術を開発する。また、SNSを利用しない高齢者を想定し、対話AI(音声対話コミュニケーションロボット)を用いて普段の生活習慣の聞き取りを行うことで、達成可能な生活習慣改善案を最適化アルゴリズムに基づき提示し、生活習慣に関するTips(アドバイス)を大規模汎用言語モデルを用いて自動生成して提示するシステムを構築する。本研究により、糖尿病をはじめとする生活習慣病のリスクが高いユーザの早期発見と早期対策に役立つテキストマイニング技術および、生活習慣改善のための気付きを与える技術を実現する(図1)。

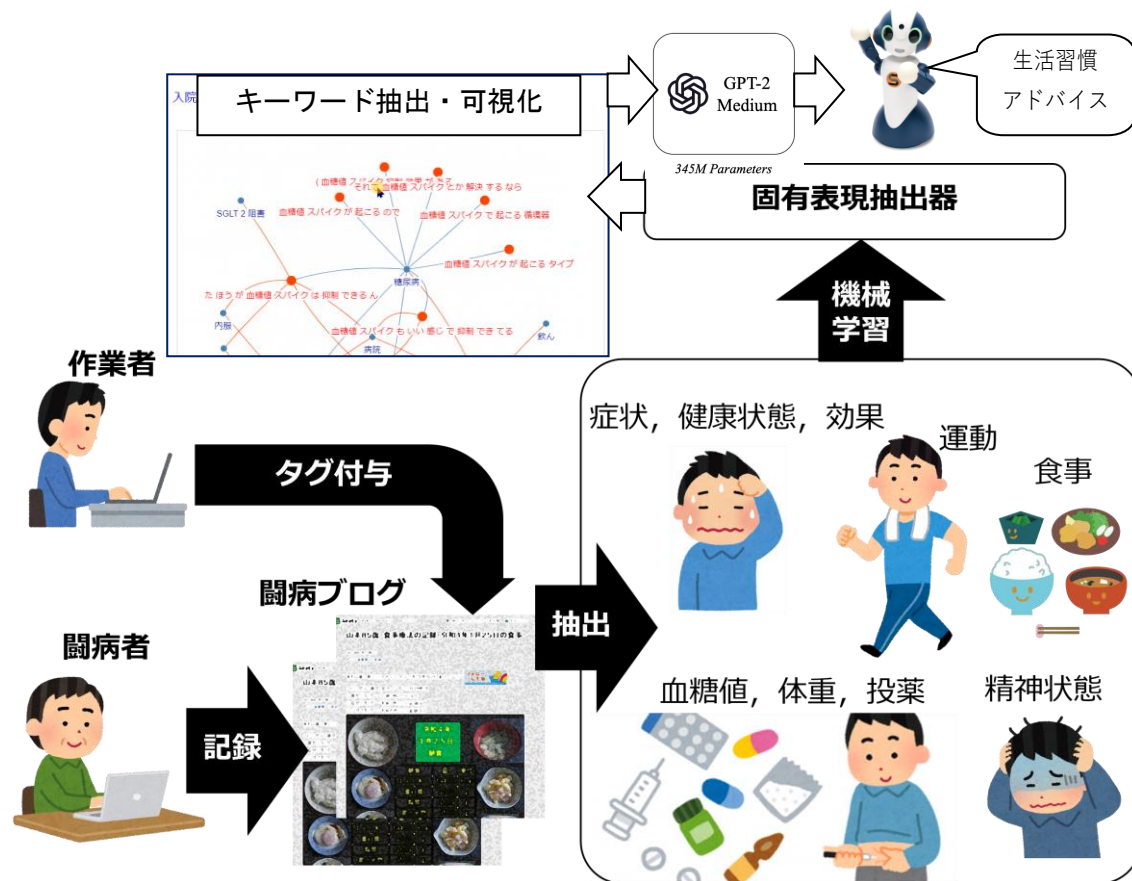


図1:研究概要

## 2 研究の目的と背景

わが国における死亡原因の約半数を生活習慣病が占めている。令和元年の死亡原因割合の調査ではがん、心疾患、脳血管疾患で50%を占めており、これらは、いずれも食事、運動、睡眠などの生活習慣に起因する生活習慣病とよばれるものである。生活習慣病の中でも国内のがん患者や心疾患患者は年々増加傾向にあり、今後のわが国における生活習慣病の対策はさらに重要になってきている。生活習慣病には自覚症状がほとんどないため、病気の進行に気づき難いという特徴がある。このことより、生活習慣病の発症を抑止するため、定期的に生活習慣改善のきっかけを与えるシステムに着目した。このシステムの構築には、人々の日々の生活習慣を把握する必要がある。生活習慣を把握する手段のひとつとして、多くの人が自身の情報を発信しているソーシャルメディアがある。その中でもテキストによる情報が多く、データ収集がしやすいTwitterに注目した。本研究ではTwitterのユーザを対象とした生活習慣改善のきっかけを与えるシステムに応用可能なツイートからの生活習慣の分析を行う。また、高齢者や多忙なユーザはソーシャルメディアを利用しなかったり、あまり頻繁に利用しない傾向にあるため、別の手段での生活習慣の取得が必要である。独居老人の場合、話し相手がないことで生活習慣悪化に対して気づきを与える人がいないことから、気づいたときには手遅れになってしまう問題がある。そこで、コミュニケーションロボットを用いた音声対話システムを利用することによって、自然な形で生活習慣の取得を行うシステムにも着目した。

本研究では、2型糖尿病の闘病者が記述した闘病ブログから生活習慣と生活習慣病の悪化や改善との因果関係についてタグ付与を行い、独自に闘病ブログコーパスを構築する。このコーパスを元に生活習慣に関連するキーワードを抽出・分類するモデルの構築を行うとともに、ソーシャルメディアから生活習慣病リスクの高いユーザの検出およびリスクの高い生活習慣についてキーワード間の関係性を元に提示する手法を提案する。さらに、音声対話ロボットを用いてユーザ個人の生活習慣を聴取し、データベースに蓄積し、これをもとに生活習慣の改善につながるアドバイスやTipsを生成して提示するシステムの構築を目的とする(図2)。

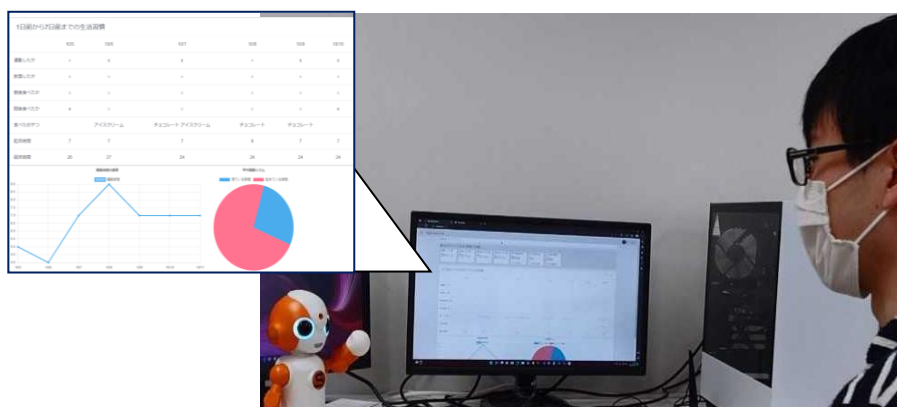


図2: 音声対話ロボットを用いた生活習慣維持支援システム

### 3 研究内容

#### (1) 闘病ブログコーパスの構築とキーワード抽出に関する研究

生活習慣病患者の生活習慣の分析のため、まず、闘病ブログの記事データ収集をTOBYOというWebサイトを用いて行った。収集対象として「2型糖尿病」に関する記事のある一定の頻度（1週間に1回以上）で投稿しているユーザの選定を行い、Pythonにより独自に作成したWebページクローラを用いて自動収集した。収集されたデータ（HTMLファイル）には出典URL、タイトル、日付、画像URL、本文テキストが含まれる（データの収集および解析の基本の部分を研究協力謝金により行い、高度な収集・解析業務については事業委託先に依頼して行った）。収集した記事データから必要な情報のみを抽出するPythonスクリプトでノイズ除去およびテキスト整形などを行った後に、brat またはdoccanoというアノテーションツールを用いて単語またはフレーズ単位でのラベル付与作業を行った（図3）。

収集されたデータの一部を用いて、深層学習モデルを用いた闘病ブログキーワードの分類に関する研究成果を、2021年7月に自然言語処理と人工知能に関する国際会議においてオンライン発表し、プロシーディングスに論文が掲載された。また、9月には、国内講演会において、研究成果の一部について発表した。2022年には、深層学習モデルに基づき構築したキーワード抽出・分類器をもとに生活習慣キーワードの抽出精度や問題点を様々な観点から分析した。

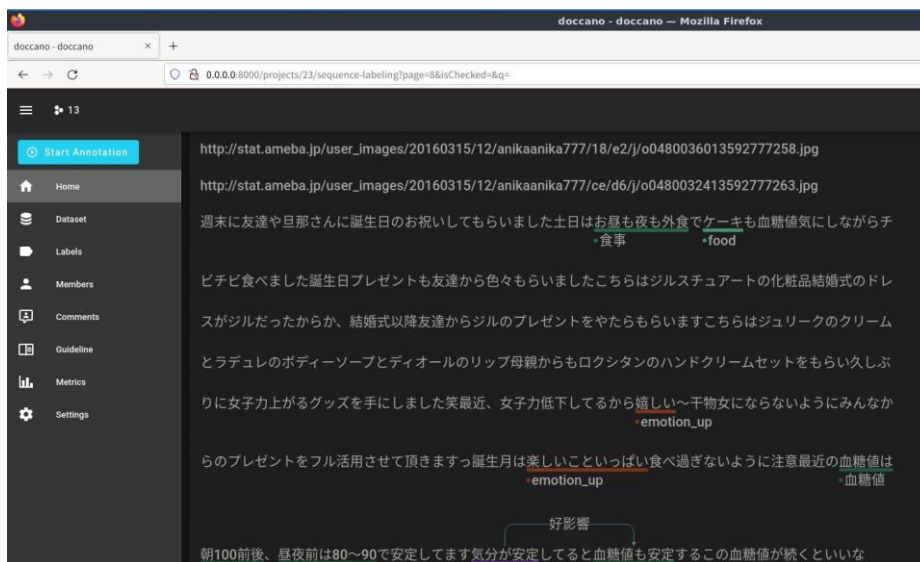


図3：doccanoによる闘病ブログに対する生活習慣キーワードのタグ付与の様子

## (2) 生活習慣を記述したツイートの収集及び分類, ユーザクラスタリング

Twitter上で健康状態や運動習慣などについての投稿があるユーザを対象に, ツイートデータの収集を行った。スパムやbotなどのツイートは対象外としてフィルタリングし, 結果として合計5213名のTwitterユーザの選定を行った。収集したデータには, 「運動」「食事」「体調」「精神状態」「その他」のラベルを半自動付与し, 運動(38297件), 食事(36255件), 体調(8936件), 精神状態(890件), その他(6669件)への付与を行った。

また, 生活習慣病のリスクの高いユーザを, 「血压」「血糖値」「体重」などといった数値の指標をもとに抽出し, 「高い」「中程度」「低い」に分けて分析を行った。具体的には, tf-idfとよばれる単語の重要度スコアを文書頻度などをもとに計算する手法と, 日本語係り受け解析を用いて, 重要な語やキーワードと共起する頻度の高い語を, 共起ネットワークを表示することで人間にわかりやすい形で提示する方法を検討した。

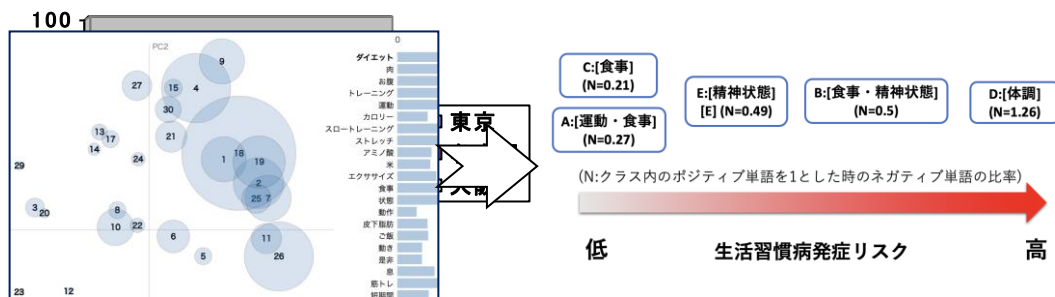


図4: ツイートユーザのクラスタリングに基づく生活習慣病発症リスクのレベル分類結果

## (3) 生活習慣維持支援システムの開発

生活習慣を聴取し, 改善案を提示するロボット対話システムを, 音声コミュニケーションロボットSotaを用いて実装した。プログラミング言語にはJavaとPythonを用いて, 具体的には以下の図5のような流れで生活習慣を聴取し, ユーザごとに生活習慣データを蓄積して生活習慣改善目標を提示するシステムを構築した。改善案の生成には, PSO(粒子群最適化)を用い, 生活習慣Tipsの生成には, 汎用言語モデルGPT-2を生活習慣改善に関するWebページデータでファインチューニングしたモデルを用いた。また, 日本語対話モデルをファインチューニングすることで, 一貫した人格(個性)を表現可能な対話モデルの構築についても研究を行った。

下記URLで本研究で開発した音声対話ロボットを用いた生活習慣維持支援システムのデモ動画を公開している。

[https://youtu.be/r1oFpb0\\_INY](https://youtu.be/r1oFpb0_INY)

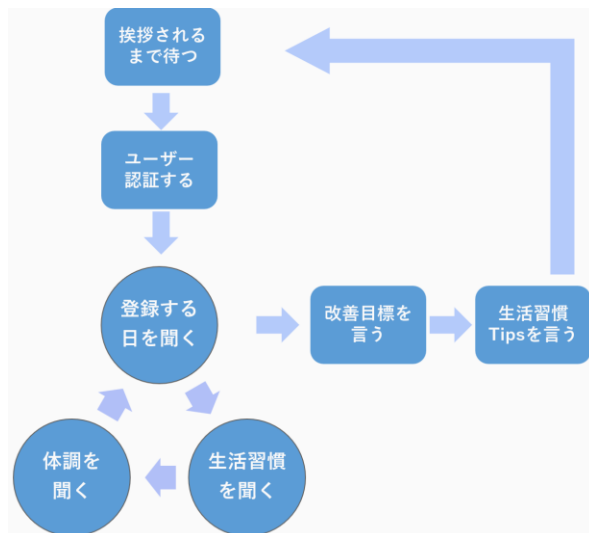


図5: 音声対話ロボットによる生活習慣聴取の流れ

#### 4 本研究が実社会にどう活かされるか—展望

本研究は、我々が今まで研究してきた技術を活用して、実社会で実用される情報処理システムを開発するものである。特に、我が国で深刻な問題となっている独居高齢者の健康管理の困難さを解消することに有用であると考えている。本研究で開発した音声対話ロボットを用いた生活習慣聴取および改善案提示の技術を、音声認識や画像認識、音声対話応答技術など、近年目まぐるしく発展を遂げている人工知能技術を導入して性能向上させていくことで、さらに応用範囲を広げることができ、生活習慣聴取および改善支援のアプリケーションやサービスがこれから一般に浸透することが期待される。

#### 5 教歴・研究歴の流れにおける今回研究の位置づけ

これまで、感性情報処理、自然言語処理およびテキストマイニングの応用研究を実施してきた中で、本研究は、それらの研究で培った知見や技術を統合して実用的なアプリケーションを開発するものである。初期段階での自覚症状に乏しい生活習慣病を早期発見するためには、まず患者自身が生活習慣病リスクについて知り、受診のきっかけを与えることが大切である。そのために普段の何気ない行動をソーシャルメディア上の投稿内容や、ロボットとの対話を通して把握し、リスクを軽減するための行動指針となる生活習慣改善案を最適化アルゴリズムにより導き出し、最新の人工知能技術を用いたアドバイス生成を行うシステムの開発が必要となる。今回の研究により、従来には無い生活習慣、健康維持支援のシステムの基盤技術が整い、システムやサービスの展開だけにとどまらず、今後の人工知能やテキストマイニングを応用した研究の発展に活かせると考えている。

## 6 本研究にかかわる知財・発表論文等

### 【発表論文】

- [1] 糖尿病患者のブログに記述された生活習慣の抽出, 宮下 翼, 松本 和幸, 吉田 稔, 西村 良太, 北 研二, 電気学会論文誌, Vol.142, No.10, 1144-1155, 2022 年(査読有)  
(DOI: 10.1541/ieejieiss.142.1144, CiNii: 1390293633437552768)
- [2] Extraction of keywords from disease-fighting blogs by fine-tuning the T5 model  
Shohei Noguchi, Kazuyuki Matsumoto, Minoru Yoshida, Ryota Nishimura and Kenji Kita,  
2022 6th International Conference on Natural Language Processing and Information  
Retrieval, Dec. 2022(査読有)
- [3] Emotion and personality estimation of utterances in a dialogue model reflecting  
personality, Daiki Mori, Kazuyuki Matsumoto, Manabu Sasayama, Minoru Yoshida and  
Kenji Kita, Proceedings of 4th International Conference on Technological Competency  
as Caring in Nursing and Health Sciences 2022, 105, Online, Dec. 2022(査読有)
- [4] Objective Analysis Techniques for Descriptive Answer Sentences in Questionnaires  
Using Natural Language Processing and Machine Learning  
Kazuyuki Matsumoto, 4TH INTERNATIONAL CONFERENCE ON TECHNOLOGICAL  
COMPETENCY AS CARING IN NURSING AND HEALTH SCIENCES 2022, Online, Dec.  
2022(Invited talk)(査読無)
- [5] Analysis of Lifestyle Habits from Weblogs of Patients with Intractable Diseases Using  
Deep Learning, Naomichi Tabuchi, Kazuyuki Matsumoto, Minoru Yoshida, Ryota  
Nishimura and Kenji Kita, Proceedings of International Conference on Electrical,  
Computer and Energy Technologies (ICECET), Jul. 2022(査読有)
- [6] Risk Analysis of Developing Lifestyle-related Diseases based on the Content of Social  
Networking Service Posts, Naomichi Tabuchi, Kazuyuki Matsumoto, Minoru Yoshida,  
Ryota Nishimura and Kenji Kita, Innovation in Medicine and Healthcare, Part of the  
Smart Innovation, Systems and Technologies book series, Vol.308, 63-72, Jun. 2022  
(査読有)
- [7] Lifestyle Analysis via a Corpus of Disease-Fighting Weblogs, Kazuyuki Matsumoto, Ryu  
Mopuaa, Minoru Yoshida and Kenji Kita, Proceedings of 2021 IEEE 2nd International  
Conference on Pattern Recognition and Machine Learning (PRML 2021), 266-272,  
Chengdu, Jul. 2021(査読有)
- [8] 音声対話ロボットを用いた健康維持支援システムの開発, 早道 広峻, 松本 和幸, 吉田 稔,  
西村 良太, 北 研二, HAI シンポジウム 2023 講演論文集, 2023 年 3 月(査読無)
- [9] 個性を反映した対話モデルにおける発話文の感情及び性格推定, 森 大樹, 松本 和幸, 篠山  
学, 吉田 稔, 北 研二, 信学技報, 34-38, 2022 年 10 月(査読無)
- [10] インタビュー対話に基づく個性を反映した対話モデルの構築, 森 大樹, 松本 和幸, 篠山  
学, 吉田 稔, 北 研二, 令和 4 年度 電気・電子・情報関係学会 四国支部連合大会講演論  
文集, 139, 2022 年 9 月(査読無)

- [11] SNS の投稿内容に基づく生活習慣病発症リスクの分析, 田淵 尚道, 松本 和幸, 吉田 稔, 西村 良太, 北 研二, 言語処理学会 第 28 回年次大会 発表論文集, 1946-1950, 2022 年 3 月(査読無)
- [12] 闘病ブログコーパスからの生活習慣の抽出, 宮下 翼, 吉田 稔, 松本 和幸, 西村 良太, 北 研二, 令和 3 年度 電気・電子・情報関係学会 四国支部連合大会 講演論文集, 187, 2021 年 9 月(査読無)

## 7 補助事業に係る成果物

### (1)補助事業により作成したもの

#### i. 闘病ブログコーパス:

<https://github.com/Kmatsu-tokudai/diabeticCorpus>

#### ii. 生活習慣維持支援システム:

[https://youtu.be/r1oFpb0\\_INY](https://youtu.be/r1oFpb0_INY)

### (2)(1)以外で当事業において作成したもの

特に無し

## 8 事業内容についての問い合わせ先

所属機関名: 徳島大学工学部(トクシマダイガクリコウガクブ)

住 所: 〒770-8506

徳島県徳島市南常三島町2-1 総合研究実験棟703号室

担 当 者: 准教授 松本 和幸(マツモト カズユキ)

担 当 部 署: 徳島大学大学院社会産業理工学研究部

(トクシマダイガクダイガクインシャカイサンギョウリコウガクケンキュウブ)

E - m a i l: [matumoto@is.tokushima-u.ac.jp](mailto:matumoto@is.tokushima-u.ac.jp)

U R L: <https://www-a2.is.tokushima-u.ac.jp/home>